Subquestion 3: Government Expenditure

Kylie King

This document explores the relationship between the approval for yearly government spending and public trust in government by political party. Specifically, it aims to answer the following question.

Subquestion 3

Is the amount of public trust in the government correlated with the amount of government spending in a year, and is that correlation different between Democratic and Republican respondents?

Data Collection

In this section, we initialize R packages and pull data downloaded from the Federal Reserve Bank of St. Louis and Pew Research.

First I'll load in the total expenditures data. This runs quarterly, but we are going to aggregate it into yearly so that we can directly compare it to the approval data

```
suppressPackageStartupMessages(library('tidyverse'))
suppressPackageStartupMessages(library('car'))
suppressPackageStartupMessages(library('lmPerm'))
```

Expenditure data from https://fred.stlouisfed.org/series/FGEXPND. Data is stored in millions of USD (ie 4000 = $4 billion) The following data was collected from the Federal Reserve Bank of St. Louis, which synthesized the data from the U.S. Bureau of Economic Analysis. The first dataset details total government expenditures by year, and each value is stored in millions of USD (ie 4000 = $4 billion).[1] The second details total Gross Domestic Product (GDP) per year.[2]

```
options(readr.show_col_types = FALSE)
#expenditures
expenditures<- read_csv('data/FGEXPND.csv')
head(expenditures)
```

```
# A tibble: 6 × 2
  observation_date FGEXPND
```

---

[1]U.S. Bureau of Economic Analysis, Federal Government: Current Expenditures [FGEXPND], retrieved from FRED, Federal Reserve Bank of St. Louis; https://fred.stlouisfed.org/series/FGEXPND

[2]U.S. Bureau of Economic Analysis, Gross Domestic Product [GDP], retrieved from FRED, Federal Reserve Bank of St. Louis; https://fred.stlouisfed.org/series/GDP

```
   <date>              <dbl>
1 1958-01-01           83.8
2 1958-04-01           88.5
3 1958-07-01           89.2
4 1958-10-01           91.0
5 1959-01-01           89.1
6 1959-04-01           90.1
```

```
#gdp
gdp<- read_csv('data/GDP.csv')
head(gdp)
```

```
# A tibble: 6 × 2
  observation_date   GDP
  <date>           <dbl>
1 1958-01-01        468.
2 1958-04-01        472.
3 1958-07-01        486.
4 1958-10-01        500.
5 1959-01-01        510.
6 1959-04-01        523.
```

The following data is a collection of public survey responses synthesized by Pew Research Center[3]
to evaluate trust in government. It is made up of the percentages of people who who say they
'trust the government to do what is right just about always/most of the time' in each survey,
divided by political party identity/leaning.

```
parties_public_trust <- read_csv('data/trust_in_government_parties.csv')
head(parties_public_trust)
```

```
# A tibble: 6 × 4
  Date      .Source `Democrat/Lean Dem` `Republican/Lean Rep`
  <chr>     <chr>                 <dbl>                 <dbl>
1 5/19/2024 PEW                      35                    11
2 6/11/2023 PEW                      25                     8
3 5/1/2022  PEW                      29                     9
4 4/11/2021 PEW                      36                     9
5 8/2/2020  PEW                      12                    28
6 4/12/2020 PEW                      18                    36
```

---

[3]Pew Research Center, National Election Studies, Gallup, ABC/Washington Post, CBS/New York Times, and CNN
surveys. Data from 2020 and later comes from Pew Research Center's online American Trends Panel; prior data
is from telephone surveys. Data collected from https://www.pewresearch.org/politics/2024/06/24/public-trust-in-
government-1958-2024/.

Data Cleaning

In this section, we will clean the data that was pulled in the previous section, to prepare it for Exploratory Data Analysis (EDA) and hypothesis testing.

Expenditure Cleaning

To streamline this data, we will aggregate all the expenditures by year (as they are separated by category). Then, we will remove the data from Q1 and Q2 of 2025, as this would show a disproportionately low amount of government expenditure in 2025, only covering half the year.

```
#group by year
expenditures$year <- year(expenditures$observation_date)
yearly_expenditures <- expenditures |> group_by(year) |>
summarise(year_sum_exp = sum(FGEXPND, na.rm = TRUE))

#removing the incomplete 2025 data
yearly_expenditures <- yearly_expenditures[-68, ]
head(yearly_expenditures)
```

```
# A tibble: 6 × 2
   year year_sum_exp
  <dbl>        <dbl>
1  1958         352.
2  1959         362.
3  1960         374.
4  1961         399.
5  1962         434.
6  1963         454.
```

GDP Cleaning

To streamline this data, we will aggregate all the gdp by year (as they are separated by quarter). We do not have to remove any 2025 data, as this particular dataset only runs through Q4 of 2024.

```
# group by year
gdp$year <- year(gdp$observation_date)
yearly_gdp <- gdp |> group_by(year) |> summarise(year_sum_gdp = sum(GDP, na.rm
= TRUE))
```

Public Trust Cleaning

This data has multiple entries for some survey years. For the purposes of this subquestion, we will take the average of each year's surveys, so that we have a single entry for each year. While this does cause some data loss, it will streamline the data for hypothesis testing. The data loss is minimized by taking the average for each year, instead of simply dropping the duplicate year values. We will also convert the percentages to proportions, which will allow better correlation between this data and the yearly government expenditure/gdp data.

```r
#ensuring uniform date format
prop_parties_public_trust <- parties_public_trust |> mutate(Date = mdy(Date))
prop_parties_public_trust <- prop_parties_public_trust |> mutate(year =
year(Date))

#aggregating by year to prevent double entries
prop_parties_public_trust <- prop_parties_public_trust |> group_by(year) |>
    summarise(
        rep_trust_prop = mean(`Republican/Lean Rep`, na.rm = TRUE) / 100,
        dem_trust_prop = mean(`Democrat/Lean Dem`, na.rm = TRUE) / 100,
        .groups = "drop")
```
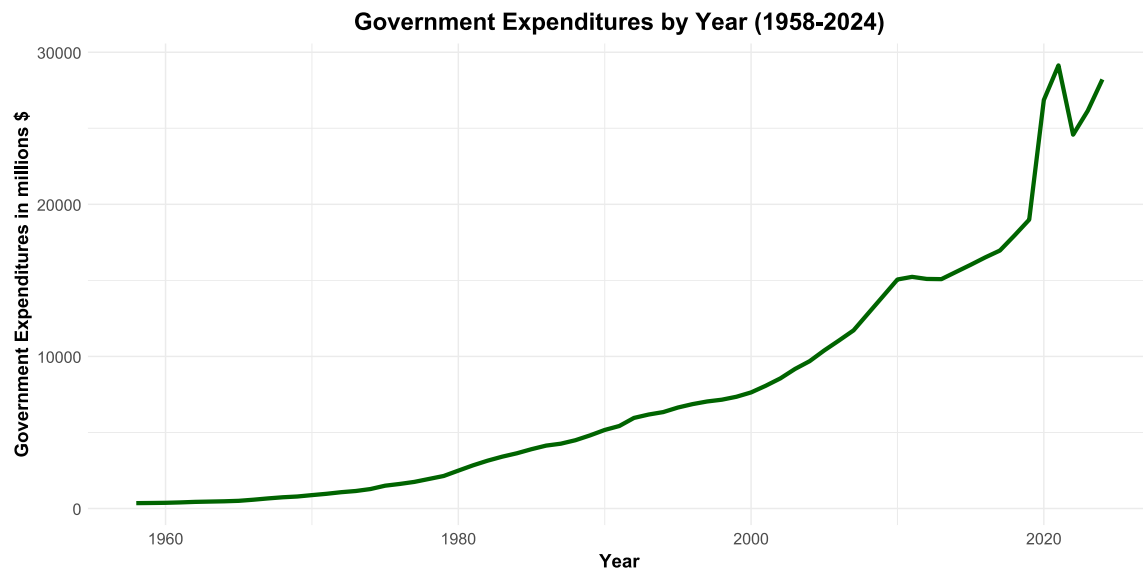
Exploratory Data Analysis

To prepare for our hypothesis tests, we will perform some EDA for each of our data. It is important to see what type of data we have, how it is distributed, and what patterns we might expect to see in the later hypothesis testing. We first initialize a theme so the charts are displayed uniformly.

```r
eda_5100 <- theme_minimal() +
  theme(
    text = element_text(color = "black"),
    plot.title = element_text(size = 18, face = "bold", hjust = 0.5),
    axis.title = element_text(size = 14, face = "bold"),
    axis.text = element_text(size = 12),
    legend.title = element_text(size = 14, face = "bold"),
    legend.text = element_text(size = 12),
    strip.text.x = element_text(size = 14, face = "bold")
  )
theme_set(eda_5100)
```

We will first examine the yearly government expenditure data from FRED.

```r
ggplot(yearly_expenditures, aes(x = year)) +
  geom_line(aes(y = year_sum_exp), size = 1.5, color='darkgreen') +
  labs(
    x = "Year",
    y = "Government Expenditures in millions $",
    title = "Government Expenditures by Year (1958-2024)"
  )
```
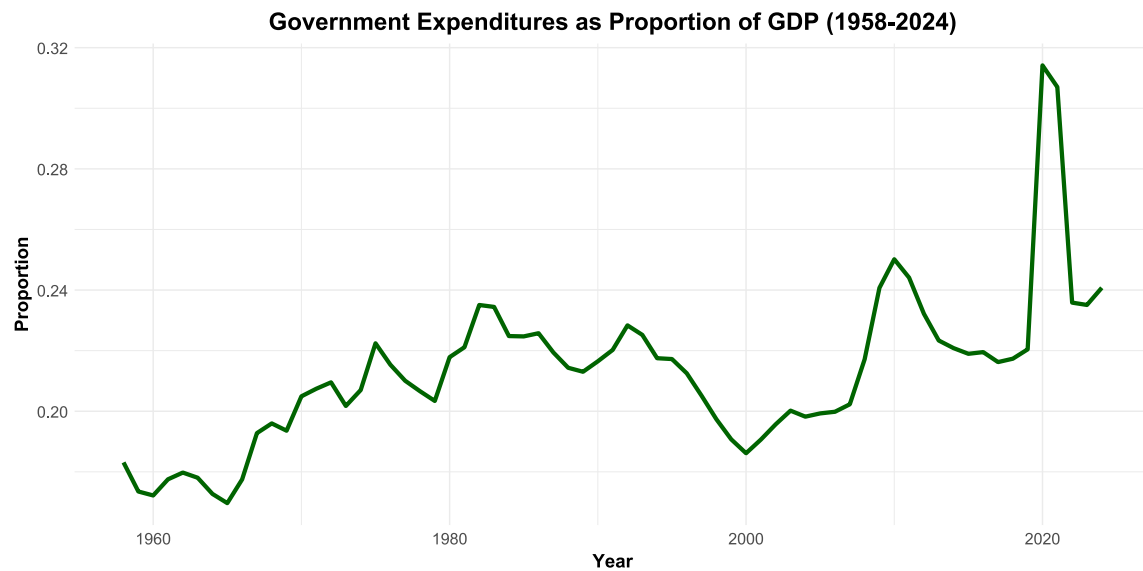
```
Warning: Using `size` aesthetic for lines was deprecated in ggplot2 3.4.0.
ℹ Please use `linewidth` instead.
```

Government Expenditures by Year (1958-2024)

As may be expected, the federal government generally spends more money each decade than the previous decade, so this unstandardized data will not offer us many insights. To standardize it, we will calculate the proportion of government expenditures to the total GDP for each year.
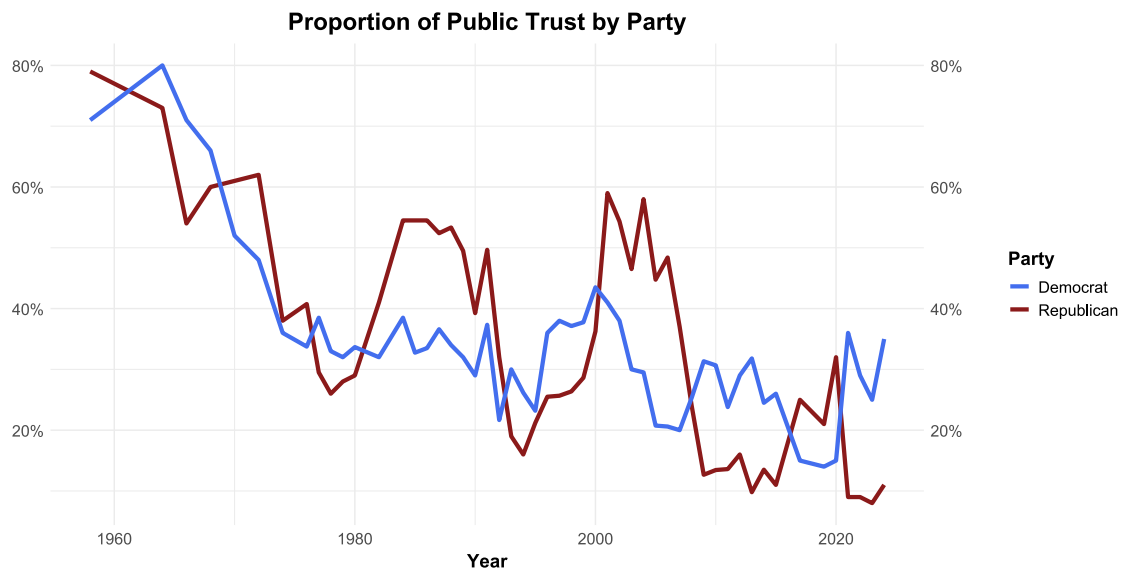
```
#now to combine our gdp and expenditures, and calculate the proportion of gdp
the expenditures are
econ_combined <- yearly_gdp |> left_join(yearly_expenditures, by = "year")
econ_combined <- econ_combined |> mutate(p_exp_of_gdp = year_sum_exp/
year_sum_gdp)

ggplot(econ_combined, aes(x = year)) +
  geom_line(aes(y = p_exp_of_gdp), size = 1.5, color='darkgreen') +
  labs(
    x = "Year",
    y = "Proportion",
    title = "Government Expenditures as Proportion of GDP (1958-2024)"
  )
```

Government Expenditures as Proportion of GDP (1958-2024)

We still see a slight increase over time, where the federal government has gradually spent more money in proportion ot its total GDP. However, this will give us a better idea of how government spending relates to public trust in the government. First, we must look at the public trust in government depending on political party preference.
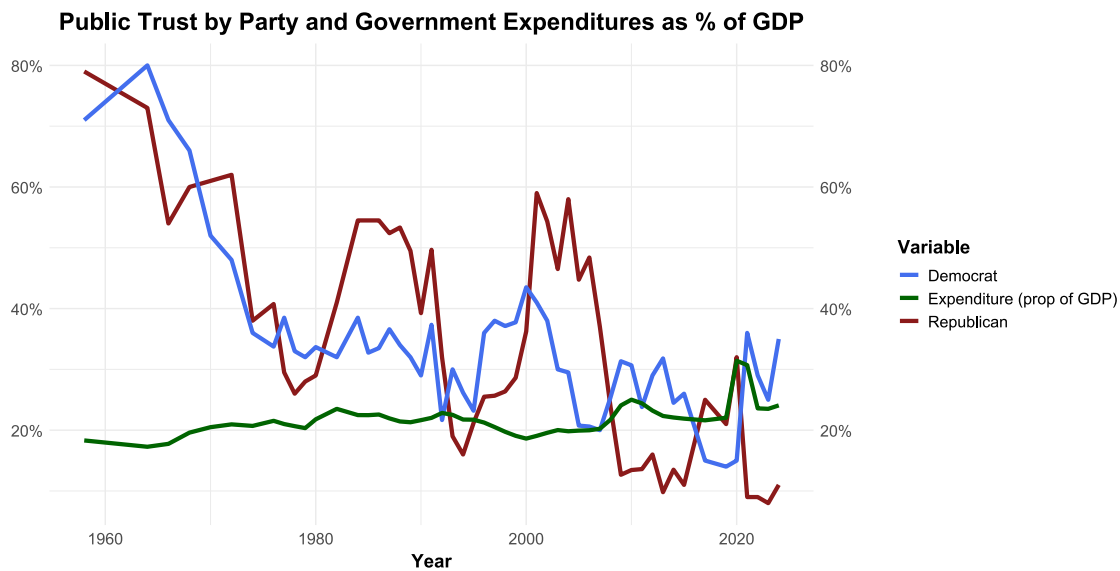
```
ggplot(prop_parties_public_trust, aes(x = year)) +
  geom_line(aes(y = rep_trust_prop, color = "Republican"), size = 1.5) +
  geom_line(aes(y = dem_trust_prop, color = "Democrat"), size = 1.5) +
  scale_y_continuous(labels = scales::percent, sec.axis = dup_axis()) +
  scale_color_manual(values = c("Republican" = "firebrick4", "Democrat" =
"royalblue2")) +
  labs(
    x = "Year",
    y = "",
    color = "Party",
    title = "Proportion of Public Trust by Party"
  )
```

Proportion of Public Trust by Party

This plot shows a gradual declne over time in the overall public trust in government, regardless of party affiliation. In addition, public trust by party appears to be inverse. When Republicans report a higher trust in the government, Democrats report lower, and vice versa. PEW's article reports that respondants are more likely to report higher trust in government when their party controls the presidency, which offers an explanation. We will now compare this standardized government expenditures data with the proportion of people who report that they 'trust the government to do what is right just about always/most of the time.'

```r
combine_exp_parties <- prop_parties_public_trust |> left_join(econ_combined,
by = "year")

ggplot(combine_exp_parties, aes(x = year)) +
  geom_line(aes(y = rep_trust_prop, color = "Republican"), size = 1.5) +
  geom_line(aes(y = dem_trust_prop, color = "Democrat"), size = 1.5) +
  geom_line(aes(y = p_exp_of_gdp, color = "Expenditure (prop of GDP)"), size =
1.5) +
  scale_y_continuous(labels = scales::percent, sec.axis = dup_axis()) +
  scale_color_manual(values = c("Republican" = "firebrick4", "Democrat" =
"royalblue2", "Expenditure (prop of GDP)" = "darkgreen")) +
  labs(
    x = "Year",
    y = "",
    color = "Variable",
    title = "Public Trust by Party and Government Expenditures as % of GDP"
  )
```

Public Trust by Party and Government Expenditures as % of GDP

There is no significant visual relationship between public trust and government expenditures, so we will have to rely on the hypothesis tests for possible connections.

Hypothesis Testing

We wish to complete an ANOVA test to determine whether the relationship between government expenditure as a proportion of gdp and the percent of Pew Research respondents who 'trust the government to do what is right just about always/most of the time' differs by political party leaning. To run an ANOVA-type test on covariance between the variables, we will run a factorial ANOVA with a continuous covariate with interaction, also known as an ANCOVA (Analysis of Covariance) test. First we must prepare the data and check assumptions.

Data Preparation

We need to make the distinct categories of political party, and combine it with the yearly government expenditure as a proportion of gdp.

```
pivoted_df <- combine_exp_parties |>
  select(year, rep_trust_prop, dem_trust_prop, p_exp_of_gdp) |>
  pivot_longer(
    cols = c(rep_trust_prop, dem_trust_prop),
    names_to = "party",
    values_to = "approval"
  ) |>
  mutate(party = ifelse(party == "rep_trust_prop", "Republican", "Democrat"))

print(pivoted_df)
```

```
# A tibble: 104 × 4
    year p_exp_of_gdp party       approval
   <dbl>        <dbl> <chr>          <dbl>
 1  1958        0.183 Republican      0.79
 2  1958        0.183 Democrat        0.71
 3  1964        0.173 Republican      0.73
 4  1964        0.173 Democrat        0.8
 5  1966        0.177 Republican      0.54
 6  1966        0.177 Democrat        0.71
 7  1968        0.196 Republican      0.6
 8  1968        0.196 Democrat        0.66
 9  1970        0.205 Republican      0.61
10  1970        0.205 Democrat        0.52
# i 94 more rows
```
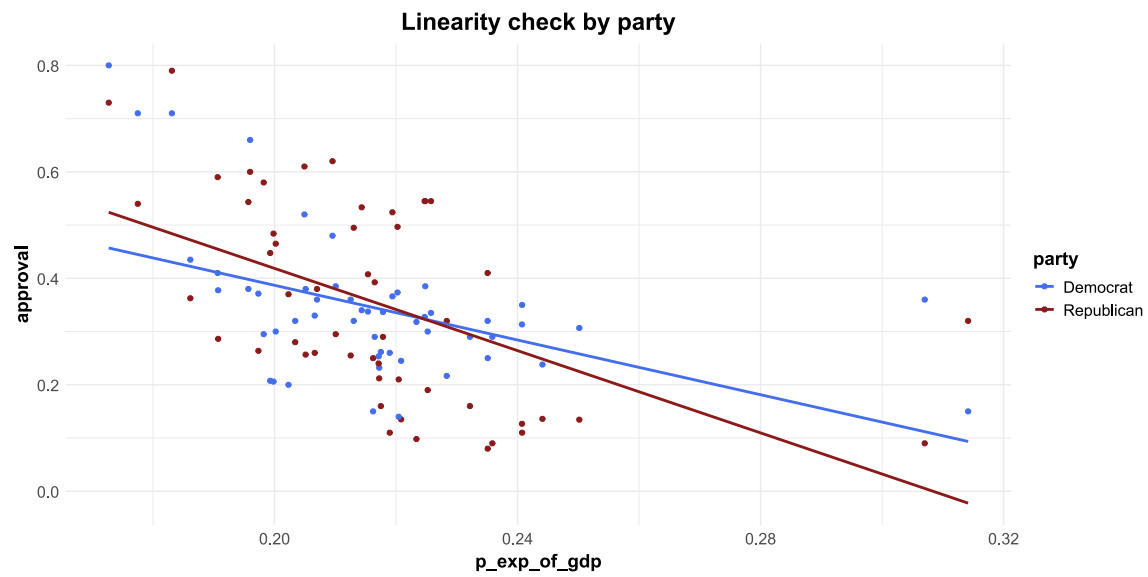
Assumptions

Next we must check our assumptions to ensure we can run a standard ANCOVA test.
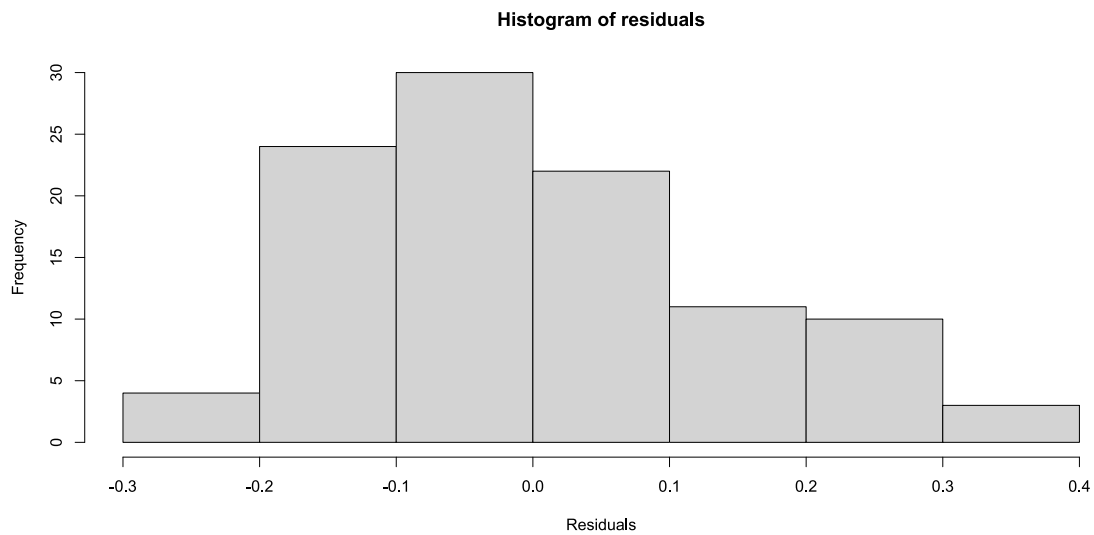
```r
#fitting the model
model <- lm(approval ~ p_exp_of_gdp * party, data = pivoted_df)

#doing a linearity check
ggplot(pivoted_df, aes(x = p_exp_of_gdp, y = approval, color = party)) +
  geom_point() +
  geom_smooth(method = "lm", se = FALSE) +
  labs(title = "Linearity check by party") +
    scale_color_manual(values = c("royalblue2", "firebrick4"))
```
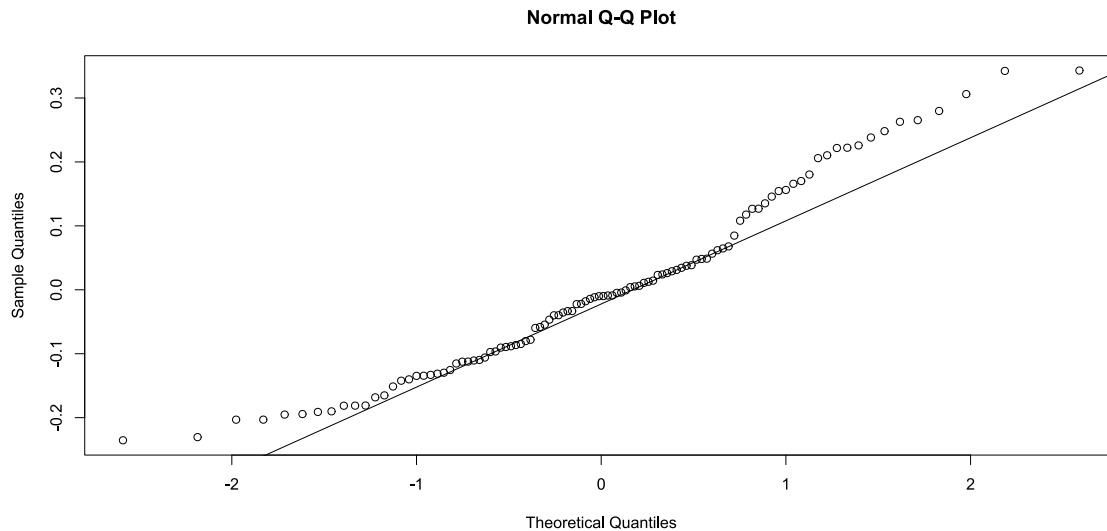
```
`geom_smooth()` using formula = 'y ~ x'
```

Linearity check by party


```
#and checking the residuals for normality: These appear normally distributed!
res <- residuals(model)
hist(res, main = "Histogram of residuals", xlab = "Residuals")
```

Histogram of residuals


```
qqnorm(res); qqline(res)
```

Normal Q-Q Plot


```r
# that deviation on the far right makes it not normally distributed
shapiro.test(res)
```

```
    Shapiro-Wilk normality test

data:  res
W = 0.96095, p-value = 0.00373
```

```r
#checking for homogenity of variance
leveneTest(res ~ pivoted_df$party)
```

```
Warning in leveneTest.default(y = y, group = group, ...): group coerced to
factor.
```

```
Levene's Test for Homogeneity of Variance (center = median)
      Df F value    Pr(>F)
group   1  14.154 0.0002814 ***
      102
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Assumptions for the factorial ANOVA with a continuous covariate with interaction, aka ANCOVA (Analysis of Covariance):

- **Independence**: The public trust surveys would have random respondands, so we can assumpe independence, and the GDP and government expenditure of one year does not necessarily directly affect the expenditures the next year.
- **Normality of Residuals**: The Shapiro-Wilk test failed, so the data is not distributed normally
- **Equality of Variances**: The levene test failed, so there is no equality of variances

We fail two of the assumptions, so we cannot continue with the parametric ANOVA test. We are going to instead do a permutation ANCOVA model. (which i also need to explain in greater detail)

Hypotheses

There are three sets of hypotheses we can extract from the permutation ANCOVA model.

Interaction: The effect of government expenditure on approval by party

Null Hypothesis: The effect of government expenditure on approval is the same for Republicans and Democrats.

Alternative Hypothesis: The effect of government expenditure on approval differs between Republicans and Democrats.

Effect of Expenditure:

Null Hypothesis: Government expenditure has no effect on approval overall.

Alternative Hypothesis: Government expenditure does affect approval overall.

Effect of Party:

Null Hypothesis: Average approval is the same for Republicans and Democrats.

Alternative Hypothesis: Average approval differs between parties.

Alpha: 0.05

We will now run the permutation ANCOVA test.

```
perm_model <- lmp(approval ~ p_exp_of_gdp * party, data = pivoted_df,
perm="Prob")
```

```
[1] "Settings:  unique SS : numeric variables centered"
```

```
summary(perm_model)
```

```
Call:
lmp(formula = approval ~ p_exp_of_gdp * party, data = pivoted_df,
    perm = "Prob")

Residuals:
```

```
      Min       1Q    Median       3Q       Max
-0.235388 -0.109903 -0.009929  0.065494  0.342998

Coefficients:
                   Estimate Iter Pr(Prob)
p_exp_of_gdp       -3.216448 5000   <2e-16 ***
party1             -0.004925   51    0.725
p_exp_of_gdp:party1 0.646640  300    0.250
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.1414 on 100 degrees of freedom
Multiple R-Squared: 0.2618, Adjusted R-squared: 0.2397
F-statistic: 11.82 on 3 and 100 DF,  p-value: 1.081e-06
```

Results and Interpretation

Residuals: These are relatively symmetrical and centered around 0, meaning we do not have any large outliers

Coefficients:

- There is a statistically significant negative effect of expenditure on overall approval. As expenditure increases, approval for both parties decreases
- There is no significant difference in average approval between parties.
- The interaction between expenditure and party approval is roughly the same for both parties.

Multiple R-Squared: 37.58% of variance in approval is explained by expenditure, party, and their interaction. F-statistic and small p-value: The model is significant

In context of the above hypotheses:

- Interaction: We fail to reject the null hypothesis that the effect of government expenditure on approval is the same for Republicans and Democrats
- Effect of Expenditure: We **reject** the null hypothesis that government expenditure has no effect on approval overall
- Effect of Party: We fail to reject the null hypothesis that average approval is the same for Republicans and Democrats

In the context of our original sub-question, "Is the amount of public trust in the government correlated with the amount of government spending in a year, and is that correlation different between Democratic and Republican respondents?", we find that yes, the amount of public trust in the government has a negative correlation with the amount of government spending in a year (as a proportion of the total GDP), but that the correlation is not significantly different between Republican and Democratic-leaning respondants.